

determining from the structure of the parse structure and the syntactic elements a grammatical role for each meaningful term;

determining an additional grammatical role for at least one of the meaningful terms, such that the at least one meaningful term is associated with at least two different grammatical roles; and

storing in an enhanced data representation data structure a representation of each association between a meaningful term and its determined grammatical roles, in a manner that indicates a grammatical relationship between a plurality of the meaningful terms and such that at least one meaningful term is associated with a plurality of grammatical relationships.

3. (New) The method of claim 2 wherein heuristics are used to determine the additional grammatical role for the at least one of the meaningful terms.

A 1  
4. (New) The method of claim 3 wherein a meaningful term is associated with a verb modifier as the determined grammatical role and is associated with an object as the additional grammatical role.

5. (New) The method of claim 3 wherein a meaningful term is associated with a verb modifier as the determined grammatical role and is associated with a subject as the additional grammatical role.

6. (New) The method of claim 3 wherein a meaningful term is associated with a verb modifier as the determined grammatical role and is associated with a verb as the additional grammatical role.

7. (New) The method of claim 3 wherein a meaningful term is associated with a subject as the determined grammatical role and is associated with an object as the additional grammatical role.

8. (New) The method of claim 3 wherein a meaningful term is associated with a object as the determined grammatical role and is associated with a subject as the additional grammatical role.

9. (New) The method of claim 3 wherein a meaningful term is associated with a noun modifier as the determined grammatical role and is associated with a subject as the additional grammatical role.

10. (New) The method of claim 3 wherein a meaningful term is associated with a noun modifier as the determined grammatical role and is associated with an object as the additional grammatical role.

A / 11. (New) The method of claim 2 wherein the determined additional grammatical role is a part of grammar that is not implied by the position of the at least one meaningful term relative to the structure of the sentence.

12. (New) The method of claim 2 wherein heuristics are used to determine which grammatical relationships are to be stored in the enhanced data representation data structure.

13. (New) The method of claim 2 wherein the determining the grammatical role for each meaningful term and the determining of the additional grammatical role for at least one of the meaningful terms yields a plurality of grammatical relationships between meaningful terms that are identical.

14. (New) The method of claim 2 wherein the determining of a grammatical role for each meaningful term includes determining whether the term is at least one of a subject, object, verb, part of a prepositional phrase, noun modifier, and verb modifier.

15. (New) The method of claim 2 wherein the document is part of a corpus of heterogeneous documents.

16. (New) The method of claim 2 wherein the enhanced data representation data structure is used to index a corpus of documents.

17. (New) The method of claim 2 wherein the enhanced data representation data structure is used to execute a query against objects in a corpus of documents.

18. (New) The method of claim 17 wherein results are returned that satisfy the query when an object in the corpus contains similar terms associated with similar grammatical roles to the terms and their associated roles as stored in the enhanced data representation.

A 1  
19. (New) The method of claim 18 wherein the objects in the corpus are sentences and indications of sentences that satisfy the query are returned.

20. (New) The method of claim 18, further comprising returning indications of documents that contain similar terms to those found in at least one indicated sentence.

21. (New) The method of claim 18, further comprising returning indications of documents that contain similar terms to those found in at least one indicated document.

22. (New) The method of claim 17 wherein terms that are associated with designated grammatical roles are returned for each object in the corpus that contains similar terms associated with similar grammatical roles to the terms and associated roles of designated relationships from the enhanced data representation data structure.

23. (New) The method of claim 17 further comprising adding additional grammatical relationships to the enhanced data representation data structure to be used to execute a query against objects in a corpus of documents.

24. (New) The method of claim 23 wherein at least one of entailed verbs and related verbs are used to add additional grammatical relationships.

25. (New) The method of claim 17 wherein weighted results are returned that satisfy the query.

26. (New) A computer-readable memory medium containing instructions for controlling a computer processor to transform a document of a data set into a canonical representation, the document having a plurality of sentences, each sentence having a plurality of terms, by:

for each sentence,

AI parsing the sentence to generate a parse structure having a plurality of syntactic elements;

determining a set of meaningful terms of the sentence from the syntactic elements;

determining from the structure of the parse structure and the syntactic elements a grammatical role for each meaningful term;

determining an additional grammatical role for at least one of the meaningful terms, such that the at least one meaningful term is associated with at least two different grammatical roles; and

storing in an enhanced data representation data structure a representation of each association between a meaningful term and its determined grammatical roles, in a manner that indicates a grammatical relationship between a plurality of the meaningful terms and such that at least one meaningful term is associated with a plurality of grammatical relationships.

27. (New) A syntactic query engine for transforming a document of a data set into a canonical representation, the document having a plurality of sentences, each sentence having a plurality of terms, comprising:

parser that is structured to decompose each sentence to generate a parse structure for the sentence having a plurality of syntactic elements; and

postprocessor that is structured to

receive from the parser the parse structure of the sentence;

determine a set of meaningful terms of the sentence from the syntactic elements;

determine from the structure of the parse structure and the syntactic elements a grammatical role for each meaningful term;

determine an additional grammatical role for at least one of the meaningful terms, such that the at least one meaningful term is associated with at least two different grammatical roles; and

store, in an enhanced data representation data structure, a representation of each association between a meaningful term and its determined grammatical roles, in a manner that indicates a grammatical relationship between a plurality of the meaningful terms and such that at least one meaningful term is associated with a plurality of grammatical relationships.

28. (New) The query engine of claim 27 wherein the postprocessor uses heuristics to determine the additional grammatical role for the at least one of the meaningful terms.

29. (New) The query engine of claim 28 wherein the postprocessor associates a meaningful term with a verb modifier as the determined grammatical role and with an object as the additional grammatical role.

30. (New) The query engine of claim 28 wherein the postprocessor associates a meaningful term with a verb modifier as the determined grammatical role and with a subject as the additional grammatical role.

31. (New) The query engine of claim 28 wherein the postprocessor associates a meaningful term with a verb modifier as the determined grammatical role and with a verb as the additional grammatical role.

32. (New) The query engine of claim 28 wherein the postprocessor associates a meaningful term with a subject as the determined grammatical role and with an object as the additional grammatical role.

33. (New) The query engine of claim 28 wherein the postprocessor associates a meaningful term with a object as the determined grammatical role and with a subject as the additional grammatical role.

34. (New) The query engine of claim 28 wherein the postprocessor associates a meaningful term with a noun modifier as the determined grammatical role and with a subject as the additional grammatical role.

A 1  
35. (New) The query engine of claim 28 wherein the postprocessor associates a meaningful term with a noun modifier as the determined grammatical role and with an object as the additional grammatical role.

36. (New) The query engine of claim 27 wherein the determined additional grammatical role is a part of grammar that is not implied by the position of the at least one meaningful term relative to the structure of the sentence.

37. (New) The query engine of claim 27 wherein the postprocessor uses heuristics to determine which grammatical relationships are to be stored in the enhanced data representation data structure.

38. (New) The query engine of claim 27 wherein the determining the grammatical role for each meaningful term and the determining of the additional grammatical role for at least one of the meaningful terms yields a plurality of grammatical relationships between meaningful terms that are identical.

39. (New) The query engine of claim 27 wherein the determining of a grammatical role for each meaningful term includes determining whether the term is at least one of a subject, object, verb, part of a prepositional phrase, noun modifier, and verb modifier.

40. (New) The query engine of claim 27 wherein the document is part of a corpus of heterogeneous documents.

41. (New) The query engine of claim 27 wherein the enhanced data representation data structure is used to index a corpus of documents.

A 1 42. (New) The query engine of claim 27, further comprising a query processor that uses the enhanced data representation data structure to execute a query against objects in a corpus of documents.

43. (New) The query engine of claim 42 wherein the query processor returns results that satisfy the query when an object in the corpus contains similar terms associated with similar grammatical roles to the terms and their associated roles as stored in the enhanced data representation.

44. (New) The query engine of claim 43 wherein the objects in the corpus are sentences and the query processor returns indications of sentences that satisfy the query.

45. (New) The query engine of claim 43 wherein the query processor returns indications of documents that contain similar terms to those found in at least one indicated sentence.

46. (New) The query engine of claim 43 wherein the query processor returns indications of documents that contain similar terms to those found in at least one indicated document.

47. (New) The query engine of claim 42 wherein the query processor returns terms that are associated with designated grammatical roles for each object in the corpus that contains similar terms associated with similar grammatical roles to the terms and associated roles of designated relationships from the enhanced data representation data structure.

48. (New) The query engine of claim 42 wherein the query processor adds additional grammatical relationships to the enhanced data representation data structure to be used to execute a query against objects in a corpus of documents.

49. (New) The query engine of claim 42 wherein the query processor returns weighted results that satisfy the query.

A / 50. (New) A method in a computer system for transforming a document of a data set into a canonical representation, the document having a plurality of sentences, each sentence having a plurality of terms, comprising:

for each sentence,

parsing the sentence to generate a parse structure having a plurality of syntactic elements;

determining a set of meaningful terms of the sentence from these syntactic elements;

determining from the structure of the parse structure and the syntactic elements a grammatical role for each meaningful term, wherein at least one of the grammatical roles for a meaningful term is at least one of a verb modifier of a prepositional phrase and a noun modifier of a noun phrase; and

storing in an enhanced data representation data structure a representation of each meaningful term associated with its determined grammatical role, in a manner that indicates a grammatical relationship between a plurality of the meaningful units.

51. (New) The method of claim 50, further comprising storing the full grammar of the sentence.



52. (New) The method of claim 50, further comprising, when it is determined that a noun modifier grammatical role is associated with one of the meaningful terms, associating the one of the meaningful terms with a subject grammatical role, thereby indicating that the one of the meaningful terms is to be stored also as a subject of the sentence.

53. (New) The method of claim 50, further comprising, when it is determined that a noun modifier grammatical role is associated with one of the meaningful terms, associating the one of the meaningful terms with an object grammatical role, thereby indicating that the one of the meaningful terms is to be stored also as an object of the sentence.

AI 54. (New) The method of claim 50, further comprising, when it is determined that a verb modifier of a prepositional phrase is a grammatical role associated with one of the meaningful terms, associating the one of the meaningful terms with an object grammatical role, thereby indicating that the one of the meaningful terms is to be stored also as an object of the sentence.

55. (New) The method of claim 50 wherein heuristics are used to determine which grammatical relationships are to be stored in the enhanced data representation data structure.

56. (New) The method of claim 50 wherein a plurality of grammatical relationships between meaningful terms that are identical are stored in the enhanced data representation data structure.

57. (New) The method of claim 50 wherein the document is part of a corpus of heterogeneous documents.

58. (New) The method of claim 50 wherein the enhanced data representation data structure is used to index a corpus of documents.

59. (New) The method of claim 50 wherein the enhanced data representation data structure is used to execute a query against objects in a corpus of documents.

60. (New) The method of claim 59 wherein results are returned that satisfy the query when an object in the corpus contains similar terms associated with similar grammatical roles to the terms and their associated roles as stored in the enhanced data representation.

61. (New) The method of claim 60 wherein the objects in the corpus are sentences and indications of sentences that satisfy the query are returned.

A 1  
62. (New) The method of claim 60, further comprising returning indications of documents that contain similar terms to those found in at least one indicated sentence.

63. (New) The method of claim 60, further comprising returning indications of documents that contain similar terms to those found in at least one indicated document.

64. (New) The method of claim 59 wherein terms that are associated with designated grammatical roles are returned for each object in the corpus that contains similar terms associated with similar grammatical roles to the terms and associated roles of designated relationships from the enhanced data representation data structure.

65. (New) The method of claim 59 further comprising adding additional grammatical relationships to the enhanced data representation data structure to be used to execute a query against objects in a corpus of documents.

66. (New) The method of claim 65 wherein at least one of entailed verbs and related verbs are used to add additional grammatical relationships.

67. (New) The method of claim 59 wherein weighted results are returned that satisfy the query.

68. (New) A computer-readable memory medium containing instructions for controlling a computer processor to transform a document of a data set into a canonical representation, the document having a plurality of sentences, each sentence having a plurality of terms, by:

for each sentence,

parsing the sentence to generate a parse structure having a plurality of syntactic elements;

determining a set of meaningful terms of the sentence from these syntactic elements;

determining from the structure of the parse structure and the syntactic elements a grammatical role for each meaningful term, wherein at least one of the grammatical roles for a meaningful term is at least one of a verb modifier of a prepositional phrase and a noun modifier of a noun phrase; and

storing in an enhanced data representation data structure a representation of each meaningful term associated with its determined grammatical role, in a manner that indicates a grammatical relationship between a plurality of the meaningful units.

69. (New) A syntactic query engine for transforming a document of a data set into a canonical representation, the document having a plurality of sentences, each sentence having a plurality of terms, comprising:

parser that is structured to decompose each sentence to generate a parse structure for the sentence having a plurality of syntactic elements; and

postprocessor that is structured to

receive from the parser the parse structure of the sentence;

determine a set of meaningful terms of the sentence from the syntactic elements;

determine from the structure of the parse structure and the syntactic elements a grammatical role for each meaningful term, wherein at least one of the grammatical roles for a meaningful term is at least one of a verb modifier of a prepositional phrase and a noun modifier of a noun phrase; and

store in an enhanced data representation data structure a representation of each meaningful term associated with its determined grammatical role, in a manner that indicates a grammatical relationship between a plurality of the meaningful units.

70. (New) The query engine of claim 69 wherein the postprocessor stores the full grammar of the sentence.

71. (New) The query engine of claim 69 wherein the postprocessor, when it is determined that a noun modifier grammatical role is associated with one of the meaningful terms, is further structured to associate the one of the meaningful terms with a subject grammatical role, thereby indicating that the one of the meaningful terms is to be stored also as a subject of the sentence.

72. (New) The query engine of claim 69 wherein the postprocessor, when it is determined that a noun modifier grammatical role is associated with one of the meaningful terms, is further structured to associate the one of the meaningful terms with an object grammatical role, thereby indicating that the one of the meaningful terms is to be stored also as an object of the sentence.

73. (New) The query engine of claim 69 wherein the postprocessor, when it is determined that a verb modifier of a prepositional phrase is a grammatical role associated with one of the meaningful terms, is further structured to associate the one of the meaningful terms with an object grammatical role, thereby indicating that the one of the meaningful terms is to be stored also as an object of the sentence.

74. (New) The query engine of claim 69 wherein the document is part of a corpus of heterogeneous documents.

75. (New) The query engine of claim 69 wherein the enhanced data representation data structure is used to index a corpus of documents.

76. (New) The query engine of claim 69 wherein the enhanced data representation data structure is used to execute a query against objects in a corpus of documents.

77. (New) The query engine of claim 76 wherein the objects in the corpus are sentences and the query processor returns indications of sentences that satisfy the query.

78. (New) The query engine of claim 76 wherein the query processor returns indications of documents that contain similar terms to those found in at least one indicated sentence.

A1 79. (New) The query engine of claim 76 wherein the query processor returns indications of documents that contain similar terms to those found in at least one indicated document.

80. (New) The query engine of claim 76 wherein the query processor adds additional grammatical relationships to the enhanced data representation data structure to be used to execute a query against objects in a corpus of documents.

81. (New) A method in a computer system for storing a normalized data structure representing a document of a data set, the document having a plurality of sentences, each sentence having a plurality of terms, comprising:

for each sentence,

determining a set of meaningful terms of the sentence and at least one grammatical role for each meaningful term; and

storing sets of grammatical relationships between a plurality of meaningful terms based upon the determined grammatical role of each meaningful term relative to a meaningful term that is being used as a governing verb, wherein, for each meaningful term that is being used as a governing verb, the normalized data structure contains a set of meaningful terms that are subjects relative to the governing verb, a set of meaningful terms that are objects relative to the governing verb, and at least one of a set of meaningful terms that are verb modifiers of

prepositional phrases that contain the governing verb and a set of meaningful terms that are noun modifiers of noun phrases that relate to the governing verb.

82. (New) The method of claim 81, further comprising storing meaningful terms that correspond to a designated attribute.

83. (New) The method of claim 82 wherein the designated attribute is at least one of country name, date, money, amount, number, location, person, corporate name, and organization.

A 1 84. (New) A data processing system comprising a computer processor and a memory, the memory containing structured data that stores a normalized representation of sentence data, the structured data being manipulated by the computer processor under the control of program code and stored in the memory as:

a subject table having a set of meaningful term pairs, each pair having a meaningful term that is associated with a grammatical role of a verb and a meaningful term that is associated with a grammatical role of a subject relative to the verb;

an object table having a set of meaningful term pairs, each pair having a meaningful term that is associated with a grammatical role of a verb and a meaningful term that is associated with a grammatical role of an object relative to the verb;

a representation of associations between the subject table and the object table, the representation indicating, for each meaningful term associated with the grammatical role of the verb, the meaningful terms that are associated with the grammatical role of subject relative to the verb and the meaningful terms that are associated with the grammatical role of object relative to the verb;

a preposition table having a set of meaningful term groups, each group having a meaningful term that is associated with a grammatical role of a verb, a meaningful term that is associated with a grammatical role of a preposition relative to the verb, and a meaningful term that is associated with a grammatical role of a verb modifier relative to the verb; and

a noun modifier table having a set of meaningful term pairs, each pair having a meaningful term that is associated with a grammatical role of a noun and a meaningful term that is associated with a grammatical role of an noun modifier relative to the noun.

85. (New) A computer-readable memory medium containing instructions for controlling a computer processor to store a normalized data structure representing a document of a data set, the document having a plurality of sentences, each sentence having a plurality of terms, comprising:

for each sentence,

determining a set of meaningful terms of the sentence and at least one grammatical role for each meaningful term; and

A 1 storing sets of grammatical relationships between a plurality of meaningful terms based upon the determined grammatical role of each meaningful term relative to a meaningful term that is being used as a governing verb, wherein, for each meaningful term that is being used as a governing verb, the normalized data structure contains a set of meaningful terms that are subjects relative to the governing verb, a set of meaningful terms that are objects relative to the governing verb, and at least one of a set of meaningful terms that are verb modifiers of prepositional phrases that contain the governing verb and a set of meaningful terms that are noun modifiers of noun phrases that relate to the governing verb.

86. (New) A computer system for storing a normalized data structure representing a document of a data set, the document having a plurality of sentences, each sentence having a plurality of terms, comprising:

enhanced parsing mechanism that determines a set of meaningful terms for each sentence and at least one grammatical role for each meaningful term; and

storage mechanism structured to store sets of grammatical relationships between a plurality of the determined meaningful terms based upon the determined grammatical role of each meaningful term relative to a meaningful term that is being used as a governing verb, wherein, for each meaningful term that is being used as a governing verb, the normalized data structure contains a set of meaningful terms that are subjects relative to the governing verb, a set

of meaningful terms that are objects relative to the governing verb, and at least one of a set of meaningful terms that are verb modifiers of prepositional phrases that contain the governing verb and a set of meaningful terms that are noun modifiers of noun phrases that relate to the governing verb.

87. (New) The system of claim 86, the storage mechanism further structured to store meaningful terms that correspond to a designated attribute.

88. (New) The system of claim 87 wherein the designated attribute is at least one of country name, date, money, amount, number, location, person, corporate name, and organization.

A1  
89. (New) A method in a computer system for transforming an object of a data set into a canonical representation for use in indexing the objects of the data set and in querying the data set, the object being other than a text-only document and having a plurality of units that are specified according to an object-specific grammar, comprising:

for each object,

decomposing the object to generate a parse structure having a plurality of syntactic elements;

determining a set of meaningful units of the object from these syntactic elements;

determining from the structure of the parse structure and the syntactic elements a grammatical role for each meaningful unit; and

storing in an enhanced data representation data structure a representation of each meaningful unit associated with its determined grammatical role, in a manner that indicates a grammatical relationship between a plurality of the meaningful units.

90. (New) The method of claim 89 wherein the objects are audio data and the units of objects are portions of audio data.



91. (New) The method of claim 89 wherein the objects are video data and the units of objects are portions of video data.

92. (New) The method of claim 89 wherein the objects are images and the units of objects are graphical data.

93. (New) A computer-readable memory medium containing instructions for controlling a computer processor to transform an object of a data set into a canonical representation for use in indexing the objects of the data set and in querying the data set, the object being other than a text-only document and having a plurality of units that are specified according to an object-specific grammar, by:

A |  
for each object,

decomposing the object to generate a parse structure having a plurality of syntactic elements;

determining a set of meaningful units of the object from these syntactic elements;

determining from the structure of the parse structure and the syntactic elements a grammatical role for each meaningful unit; and

storing in an enhanced data representation data structure a representation of each meaningful unit associated with its determined grammatical role, in a manner that indicates a grammatical relationship between a plurality of the meaningful units.

94. (New) A query engine in a computer system for transforming an object of a data set into a canonical representation for use in indexing the objects of the data set and in querying the data set, the object being other than a text-only document and having a plurality of units that are specified according to an object-specific grammar, comprising:

decomposition processor that is structured to decompose each object to generate a parse structure having a plurality of syntactic elements; and

postprocessor that is structured to

receive from the decomposition processor the generated parse structure;